JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 93, NO. B2, PAGES 1073–1085, FEBRUARY 10, 1988

Adaptation of Back Projection Tomography to Seismic Travel Time Problems

EUGENE HUMPHREYS[1] AND ROBERT W. CLAYTON

Seismological Laboratory, California Institute of Technology, Pasadena

A back projection method of reconstruction is adapted to invert seismic travel time data for velocity structure. Adaptations are made so that the inhomogeneous and anisotropic ray sets and the three-dimensional geometries commonly dealt with in seismic experiments can be handled with greater success. Jacobi iteration, deconvolution, and ray weighting work well in augmenting the basic back projection method to produce a well-focused image. These methods succeed by amounts that depend on the quality of the ray coverage. Also, the ability to reconstruct an accurate image when the data include moderate amounts of noise is shown to be good. Comparison of inversions produced with back projection tomography and with damped least squares indicate that the two methods are comparable in their ability to reconstruct an image of the actual structure. The back projection approach, however, is much more computer efficient. In practice, this allows for the construction of more detailed inversions.

INTRODUCTION

The theoretical basis for tomography can be traced to *Radon* [1917], who formulated the forward and inverse transform pair that now bear his name. The Radon transform can often be related to the projection of a two-dimensional scalar field, such as the projection of a slowness field along a parallel ray set to produce the travel time delays associated with the rays. One of the first significant applications of this theory to a physical problem was that of *Bracewell* [1956], who devised a method to image celestial bodies with radio signals. Seismologists also have long been using the principle of the Radon transform in the construction of "slant stacks." But it has been the medical researchers who have had the most remarkable success, and they who coined the word tomography for the high-resolution, two-dimensional "tomograph" (slice picture) through a patient. Fundamental to their success has been the discretization of the space to be imaged into cells, thus posing the problem in a manner well suited to digital computers. High resolution is achieved by dividing the space into many small cells. The key inversion algorithm employed by this method is a back projection scheme in which each ray is individually traced and the signal associated with the ray is distributed in the region along the ray path. The algorithm relies on the back projection of many rays and the superposition of the associated streaks to reconstruct an image. By itself, simple back projection produces a rather blurred image. The nature of the blurring is well understood, though, and the application of procedures specifically designed to compensate for the blurring produces a high-quality image.

It is the capability of back projection tomography to produce a highly resolved image that makes it attractive for seismic application. However, such application requires the ability to handle ray sets of poor distribution and which often fill all three spatial dimensions. These are problems carefully avoided in the medical application of the technique. In this paper, adaptations to back projection tomography are discussed that allow us to deal with these more general ray geometries.

Following a discussion of the method, we address the topics of resolution and model stability in the presence of noise. Finally, we present the results of numerical experiments designed to illustrate the various algorithms discussed below.

OVERVIEW OF BACK PROJECTION TOMOGRAPHY

A common problem in seismology is the determination of the velocity structure in a region that has been sampled with a set of rays. The usual approach in formulating the inverse problem is to divide the region to be modeled into a set of blocks. A reference velocity structure is assumed, and deviations from the expected travel times are inverted to obtain the slowness perturbations of the blocks (slowness being defined as reciprocal velocity). In practice, only the assumed velocity structure is used to guide the ray's path, thus producing an approximate formulation but one in which the slowness distribution is not dependent upon itself. This linearizes the problem.

The discrete forward problem can be written

$$t_r = \sum_b l_{rb} s_b$$

where t_r is the time delay associated with the rth ray, s_b is the slowness perturbation of the bth block, and l_{rb} is the length of the rth ray segment in the bth block. In matrix form the discrete representation can be compactly written $\mathbf{t} = \mathbf{Ls}$, where $\mathbf{L}$ is an $N \times M$ matrix for N being the number of travel time data and M being the number of model blocks. Because most blocks are not hit by any particular ray, most of the elements of $\mathbf{L}$ are zero. The least squares solution to this problem is found by solving the normal equations, $\mathbf{L^TLs} = \mathbf{L^Tt}$ [e.g., *Aki et al.*, 1977]. At this point it is desired to invert the square $M \times M$ matrix $\mathbf{L^TL}$. The i–jth element of this matrix gives a measure of how well "connected" the information is between the ith block and the jth block of the model.

Because of the potentially large size of $\mathbf{L^TL}$, its construction, storage, and direct inversion can be formidable. The attraction of the back projection method is that this is avoided. The simplest approach is to approximate the inversion by initially using only the diagonal of $\mathbf{L^TL}$, which estimates the slowness of the bth block with

$$s_b = \sum_r t_r l_{rb} \Big/ \sum_r l_{rb}^2 \qquad (1a)$$

or, in general,

$$\mathbf{s} = \mathbf{D}^{-1}\mathbf{L^Tt} \qquad (1b)$$

[1]Now at Department of Geological Sciences, University of Oregon, Eugene.

Paper number 7B6018.
0148-0227/88/007B-6018$05.00

for $\mathbf{D} = \text{diag}(\mathbf{L}^T\mathbf{L})$. An efficient procedure can be used to accomplish this: Each ray is projected back from its receiver one at a time, and for each block encountered the contributions to the sums $\sum t_r l_{rb}$ and $\sum l_{rb}^2$ are accumulated in two computer storage spaces. After all rays have been back projected, each block's slowness is estimated by taking the ratio of that block's two accumulated sums. This simple scheme, called the basic back projection reconstruction, is fast and requires comparatively little storage space. The resulting model, however, has a tendency to be strongly blurred, especially along the paths commonly taken by the rays. Two general classes of algorithms, iteration and filtering, have been developed to correct this problem [e.g., *Natterer*, 1986]. These have the effect of accounting for all of $\mathbf{L}^T\mathbf{L}$ instead of only its diagonal.

In principle, filtering is a deconvolutional scheme accomplished in either the space or wave number domain. Because the tomographic reconstruction is the linear superposition of the single-block responses, deconvolution is a valid procedure (so long as an appropriate single-block response can be found). Another approach to solving the blurring problem is by iterating on the travel time residuals (where a residual is the difference between th actual delay and the model-predicted delay for a particular ray). These topics are discussed below, but first we discuss a procedure for selectively weighting rays. We have found this procedure to improve the effectiveness of filtering and iteration when nonideal ray geometries are used.

Ray Weighting

In the medical application of tomography the experimental geometry is designed to produce ray coverage that is both isotropic, i.e., rays are distributed evenly in orientation angle, and homogeneous, i.e., the geometry of the rays sampling different blocks is the same. In seismic application, usually one or both of these properties do not hold. When one has the ideal ray geometry, the reconstruction of a single anomalous block results in model slowness simply being inversely proportional to the distance from the anomalous block. When a more typical set of rays is used, however, this simple pattern becomes distorted. (The reconstruction of an isolated unit-valued anomalous block is often called that block's point spread function). The nonuniform ray distribution often produces prominent streaks that radiate from anomalous blocks along the directions most commonly taken by the rays traversing these blocks. This is a natural result of the averaging represented by (1*a*); the anisotropy of the ray set hitting a block will produce a similarly anisotropic point spread function. By reducing the weight given rays alighed in common orientations, this effect can be reduced and the character of the point spread function can be improved. To accomplish this, each ray is weighted in inverse proportion to some measure of the ray density in that particular ray's direction. This modification can easily be accommodated by (1) with the inclusion of a weighting parameter w_{rb} in both the numerator and denominator:

$$s_b = \sum_r w_{rb} t_r l_{rb} \Big/ \sum_r w_{rb} l_{rb}^2 \qquad (2)$$

We have tried two approaches to weighting. In the first case, each ray has a weight that is inversely proportional to the number of rays within a narrow range of angle centered on the orientation angle of the ray in question, as determined by the angular distribution of the entire set of rays. This weight is the same along the entire length of the ray, and we can substitute w_r for w_{rb}. This approach is straightforward and can be rapidly employed, but it depends on the distribution of the complete ray set being a fairly faithful representation of the ray set investigating each block, that is, on the ray set being nearly homogeneous. When this is the case, the method will usually reduce the streaking problem significantly, even for strongly anisotropic ray geometries.

With our second approach, weights are determined in a three-step procedure: (1) delays associated with rays sampling each block are divided into several arbitrary orientation subsets, (2) the average for each subset is found, and (3) these separate determinations are averaged to arrive at a whole-block estimate. With this approach, a ray will have weight that is inversely proportional to the ray count of the subset to which it belongs. The homogeneity of the ray set is much less critical because this weight is determined for each block through which the ray passes, but a ray will generally possess a weight that varies along its length.

Filtering

The linear nature of (1) implies that the reconstruction resulting from a complex structure is simply the linear superposition of the reconstructions of each of the individual blocks. This linearity can be taken advantage of with the use of filtering algorithms.

When one is dealing with a two-dimensional situation having homogeneous and isotropic ray coverage, the reconstruction of an isolated anomalous block resulting from (1) will produce a $1/r$ point spread function centered on the center of the anomalous block, where, for any given block, r is the distance from the center of the anomalous block to the block under consideration. *Rowland* [1979] has shown that the functions $1/r$ (in space), and $1/|k_r|$ (in the wave number domain, where k_r is the wave number) are space-wave-number counterparts. For the two-dimensional case, multiplying the Fourier transformed slowness model by $|k_r|$ and then transforming back to the space domain will properly compensate for the $1/r$ blurring. Of course, when taking this product, the function $|k_r|$ must be truncated (or otherwise terminated) for wave numbers higher than are of interest because this function becomes increasingly large away from the origin.

In three dimensions the point spread function of a single anomalous block is $1/r^2$. It is still true, however, that multiplication by $|k_r|$ corrects for the radial blurring. This can be seen by Fourier transforming r^{-2}, which when spherical symmetry applies and r is the distance from the origin, reduces to [*Bracewell*, 1965]

$$f(k_r) = 4\pi \int_0^\infty (r^{-2}) \operatorname{sinc}(k_r r/\pi) r^2 \, dr = (2\pi^3 |k_r|)^{-1} \qquad (3)$$

The space domain offers an alternate approach to deconvolution which avoids the direct use of Fourier transforms. The space domain filter which is the Fourier inverse of $|k_r|$ weighting in the wave number domain can be found by rewriting $|k_r|$ as $(-i|k_r|/k_r)(ik_r)$. This is the application of the Hilbert transform and the radial derivative (for brevity, **hd**). In space the convolution of these two operators gives a filter with a central peak, and broad, negative side lobes (Figure 1). The application of $|k_r|$ can be viewed as a high wave number enhancer, like the first derivative, though the response is kept phaseless by the application of the Hilbert transform.

To relate filtering to the matrix formulation of the problem,

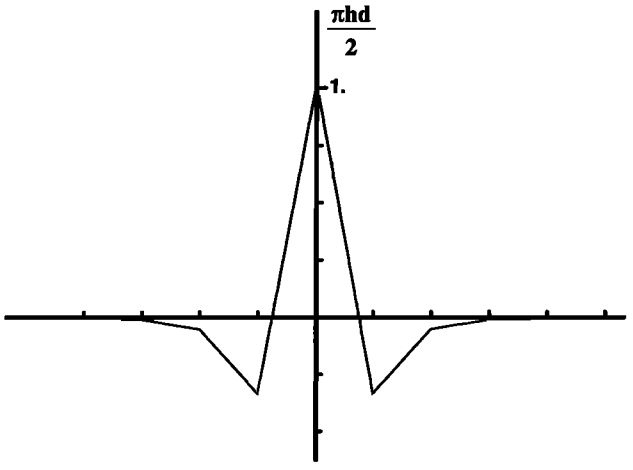

Fig. 1. The convolution of the Hilbert transform operator and the first derivative operator, **hd**. This is the space domain deconvolution operator that will correct for the model blurring that results from the basic back projection of an isotropic ray set.

we can premultiply the normal equations by $\mathbf{D}^{-1}$ to get $\mathbf{D}^{-1}[\mathbf{L}^T\mathbf{L}]\mathbf{s} = \mathbf{D}^{-1}\mathbf{L}^T\mathbf{t}$. With **D** taken to be the diagonal of $\mathbf{L}^T\mathbf{L}$, the jth column of $\mathbf{D}^{-1}[\mathbf{L}^T\mathbf{L}]$ is seen to be the jth point spread function by setting all model slowness values to zero except for the jth, which we set to unity. The left side is then the jth row of $\mathbf{D}^{-1}\mathbf{L}^T\mathbf{L}$, while the right side is the basic back projection inverse, i.e., the result of (1). Applying deconvolution to the left-hand side therefore leaves one simply with **s**, and thus deconvolution is seen to account for all of $\mathbf{L}^T\mathbf{L}$.

So far, filtering has been discussed only for the case of isotropic and homogeneous ray coverage, and the approaches have been exact. When the ray coverage is not isotropic and homogeneous, the general approaches just described may become approximate. When the ray set is anisotropic yet homogeneous, wave number deconvolution will perform properly, so long as the point spread function can be determined. But if the ray coverage varies from location to location, the point spread function for each block will be different, and wave number deconvolution may not work well. If, however, the ray coverage is not strongly heterogeneous, it has been found that deconvolving with an average point spread function works reasonably well. When using this approach, stabilizing procedures are usually applied to keep the deconvolution-produced model from becoming excessively energetic.

When the ray coverage is moderately heterogeneous, space domain filtering has proven to be a more useful approach. Since only an incomplete focusing can be accomplished, the space domain filter is approximated and, for convenience, kept of small range. This is to enable easier application and to minimize interference with the model boundaries. Toward this end, a filter **f**, of arbitrary length can be constructed to have as similar an effect as **hd** (in a least squares sense) as possible when applied to a slowness distribution **s**, i.e., minimize the energy of $((\mathbf{f} - \mathbf{hd}) * \mathbf{s})$ for a given **s**, where $*$ represents convolution. The nearest-neighbor filter is the simplest such filter. In the practical use of this filter, stabilization has been incorporated by linearly combining the output of the filter with the input. Symbolically,

$$\mathbf{s}_{new} \leftarrow a(\mathbf{s}_{old}) + b(\mathbf{f} * \mathbf{s}_{old}) \qquad (4a)$$

where a and b are least squares determined constants which minimize the squared difference between the actual delays and those predicted by the model. While the determination of **f** depends upon the existing estimated slowness distribution $\mathbf{s}_{old}$, it can be stated that since **hd** is symmetrical, **f** must also be symmetrical. We can choose the central filter coefficient to have a value of unity without loss of generality. If **f** were altered to any symmetrical nearest-neighbor filter with central weight of unity and arbitrary nearest-neighbor weight c, a and b can be adjusted so as to leave (4a) unchanged. This implies that with the use of (4a) and the nearest-neighbor representation of **hd**, **s** has no bearing on the determination of **f**. Furthermore, with appropriate choice of c, (4a) can be recognized as a linear combination of the identiy operator and the discrete Laplacian operator,

$$\mathbf{s}_{new} \leftarrow (a + b\nabla^2)\mathbf{s}_{old} \qquad (4b)$$

Iteration

The final image enhancement discussed is the application of iteration. Several approaches are possible, including the stationary back projection methods of algebraic reconstruction technique [*Herman et al.*, 1973] and simultaneous iterative reconstruction technique (SIRT) [*Gilbert*, 1972], and the so-called conjugate gradient methods [e.g., *Nolet*, 1985; *Paige and Saunders*, 1982*a*, *b*]. Below, we employ the SIRT algorithm. This method iterates on the difference between the observed delays and those predicted by the latest model, back projecting this difference, and adding the resulting correction to the existing model to form an updated version of the estimate. This process is then repeated. With respect to the matrix formulation of the problem, we can show that each iteration is equivalent to a single Jacobi iteration. To do this, rewrite the normal equations $[\mathbf{D} - (\mathbf{D} - \mathbf{L}^T\mathbf{L})]\mathbf{s} = \mathbf{L}^T\mathbf{t}$, which can be formally rearranged to $\mathbf{s} = \mathbf{D}^{-1}\mathbf{L}^T\mathbf{t} + (\mathbf{I} - \mathbf{L}^T\mathbf{L})\mathbf{s}$, where **D** is again chosen to be the diagonal of $\mathbf{L}^T\mathbf{L}$. Recognizing $\mathbf{D}^{-1}\mathbf{L}^T$ to be the back projection operator (from (1b), the iteration relation that adds the back projected residuals to the existing model can be written $\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \mathbf{D}^{-1}\mathbf{L}^T(\mathbf{t} - \mathbf{t}^{(k)})$, where $^{(k)}$ refers to the kth approximation of the solution and $\mathbf{t}^{(k)}$ are the delays predicted by the kth slowness distribution: $\mathbf{t}^{(k)} = \mathbf{L}\mathbf{s}^{(k)}$. This iterative sequence can be started by assuming that $\mathbf{s}^{(0)}$ is zero, and we find that the first iteration gives $\mathbf{s}^{(1)} = \mathbf{D}^{-1}\mathbf{L}^T\mathbf{t}$, i.e., the simple back projection inverse given by (1).

There are two questions concerning the convergence of this iterative series: Is it convergent, and if so, to what does it converge? *Ivansson* [1983] has shown that if the similarly posed iterative reconstruction algorithm of *Dines and Lytle* [1979] converges, it converges to the least squares problem. R. Comer and R. Clayton (unpublished manuscript, 1987) have shown that a general class of algorithms converge to their associated least squares solutions and have found conditions under which convergence can be guaranteed.

In general, the iteratively generated series shown above is divergent unless the update is scaled. We discuss below that with proper scaling, it converges to the least squares solution. The simplest approach to scaling is simply to multiply the update by a scaler constant that is small enough to insure convergence. More efficient procedures are available, however. Here we consider two approaches. First, we can prevent divergence by linearly combining the update with the existing slowness,

$$\mathbf{s}^{(k+1)} = a\mathbf{s}^{(k)} + b\mathbf{D}^{-1}\mathbf{L}^T(\mathbf{t} - \mathbf{t}^{(k)}) \qquad (5)$$

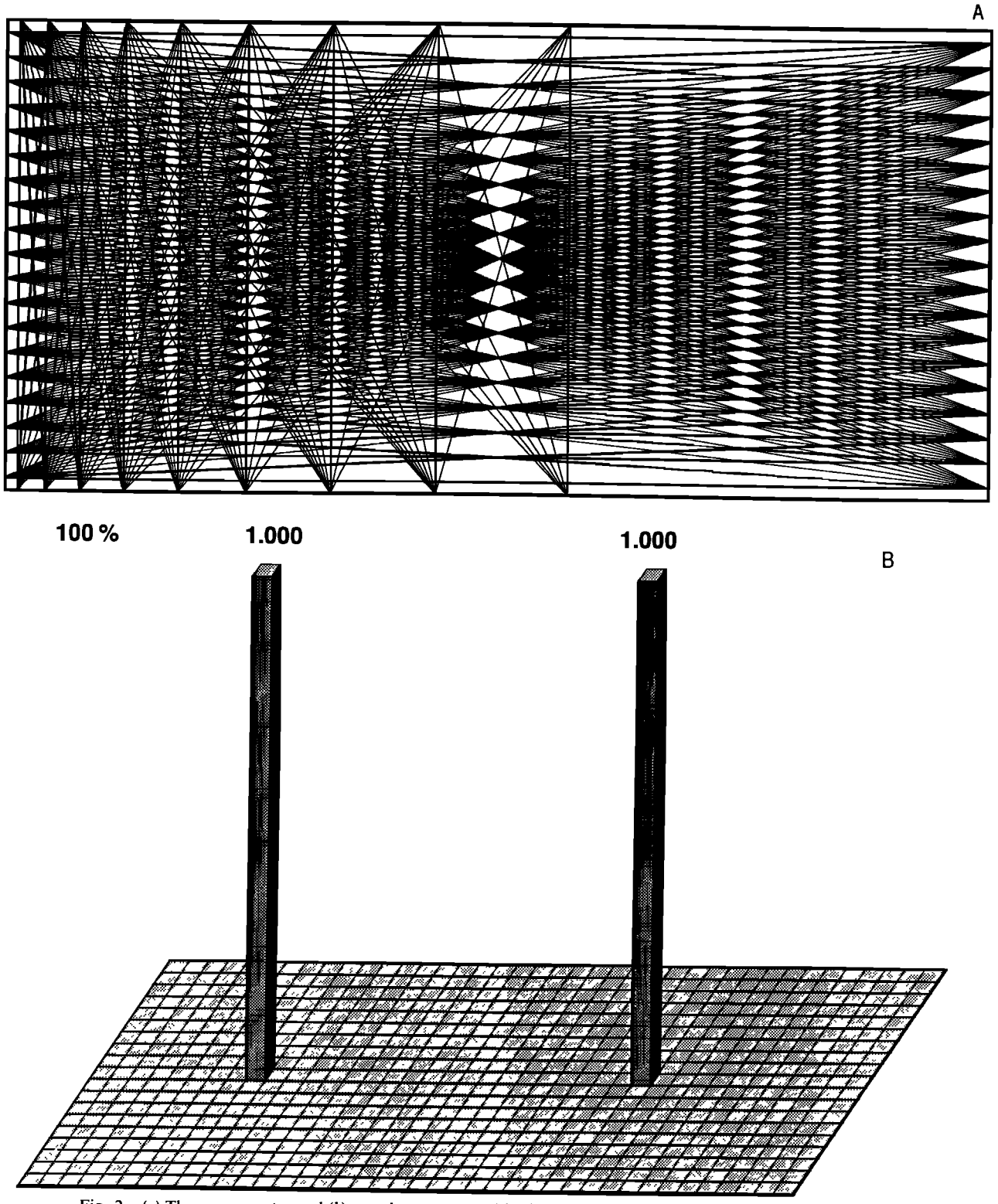

Fig. 2. (*a*) The ray geometry and (*b*) actual structure used in the generation of the following figures. In Figure 2*a* note that the ray geometry is strongly anisotropic and inhomogeneous. Figure 2*b* shows the actual structure: two isolated blocks, one located in a region of highly anisotropic ray coverage and the other in a region of roughly isotropic ray coverage. On this figure and following figures the amplitude of the major peaks are indicated above the appropriate peak. Also shown is the root-mean-squared percent of the time delays accounted for by the structure. The vertical scale has been kept constant in the following plots of this type so that they can be compared directly.

by finding constants a and b which minimize the squared model-predicted residuals, $(\mathbf{t} - \mathbf{t}^{(k+1)})^2$. This works well in practice, as is demonstrated below, but convergence properties are difficult to show. On the other hand, convergence can be guaranteed by including the $N \times N$ diagonal matrix $\mathbf{N}^{-1}$,

$$\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} + \mathbf{D}^{-1}\mathbf{L}^{\mathrm{T}}\mathbf{N}^{-1}(\mathbf{t} - \mathbf{t}^{(k)}) \tag{6}$$

for proper choice of $\mathbf{N}$. By considerations similar to those above, we can find that this relation is an iterative solution to $[\mathbf{N}^{-1/2}\mathbf{L}]^{\mathrm{T}}[\mathbf{N}^{-1/2}\mathbf{L}]\mathbf{s} = [\mathbf{N}^{-1/2}\mathbf{L}]^{\mathrm{T}}[\mathbf{N}^{-1/2}\mathbf{t}]$. This is the

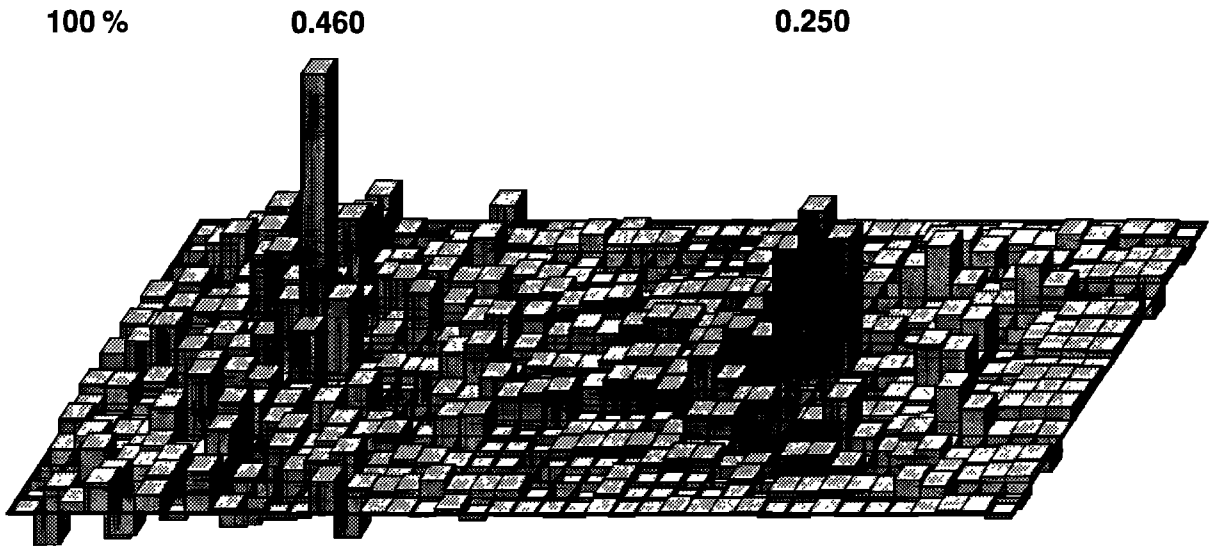

Fig. 3. The generalized inverse of the delays produced by the structure shown in Figure 2*b*. An insignificant amount of damping has been included (damping constant equal to 10^{-12}, which is 10 orders of magnitude less than the minimum significant eigenvalue shown in Figure 4). This structure accounts for virtually all of the delays, that is, within the accuracy of the computer, this structure will produce the same set of delays as the structure shown in Figure 2*b*.

"maximum likelihood" problem for data covariance matrix **N** [*Aki and Richards*, 1980], and following *Ivansson* [1983], we can show that if (6) converges, it converges to the maximum likelihood solution. If every ray segment along an individual ray is of equal length, then we can guarantee convergence by choosing $N_{rr} = n_r$ for n_r being the number of blocks hit by the rth ray (following R. Comer and R. Clayton, unpublished manuscript, 1987). If ray segments are of different lengths, convergence can still be guaranteed using $N_{rr} = n_r \times [{}^{\max}l_{rb}/{}^{\min}l_{rb}]$ for maximum and minimum ray segment lengths of the rth ray, ${}^{\max}l_{rb}$ and ${}^{\min}l_{rb}$, respectfully.

Scaling by **N** diminishes the amplitude of the update. It also gives greater relative emphasis to those rays which hit a fewer number of blocks. With λ_i being the ith eigenvalue of $\mathbf{A} = \mathbf{D}^{-1}\mathbf{L}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{L}$, the iteratively generated series can be viewed as successively "building up" the eigenvalues of $\mathbf{A}^{-1}$. At k iterations we have, instead of $1/\lambda_i$, $[1/\lambda_i] \times [1 - (1 - \lambda_i^2)^k]$ (R. Comer and R. Clayton, unpublished manuscript, 1987). A similar expression is given by *Ivansson* [1983]. We see that convergence requires all eigenvalues to be of magnitude less than 2, and if we desire the largest eigenvalues to be of greatest importance, we require all eigenvalues to be of magnitude less than 1. It is also clear that eigenvalues of small magnitude will require a greater number of iterations before they will be well represented. This is similar in effect to the use of damping with the least squares method. If the influence of small eigenvalues are desired, however, the large number of iterations needed to include these may become a problem.

Resolution and Error Estimations

If one has constructed **L**, such as when using the generalized inverse, the resolution and sensitivity of the model to noise can be simply and directly estimated [*Wiggins*, 1972]. Resolution kernels can be constructed, and the model covariance matrix can be used to infer the sensitivity of the model to noise. When dealing with a detailed model, there is the problem that **L**, $\mathbf{L}^{\mathrm{T}}\mathbf{L}$, and $\mathbf{L}\mathbf{L}^{\mathrm{T}}$ are of very large proportions and may be difficult to store in a computer. As mentioned above, when using back projection tomography, we never actually construct any of these matrices. But without the availability of this matrix, the question arises as to how one handles the estimations of resolution and the sensitivity to noise.

Resolution is determined, when using the generalized inverse, with the resolution matrix, $\mathbf{R} = \mathbf{L}^{\mathrm{T}}(\mathbf{L}\mathbf{L}^{\mathrm{T}})^{-1}\mathbf{L}$, and $\mathbf{s} = \mathbf{R}\hat{\mathbf{s}}$ [*Backus and Gilbert*, 1968; *Wiggins*, 1972], where the jth row of **R** gives the weighting coefficients (jth resolution kernel) applied to the "actual" slowness values ($\hat{\mathbf{s}}$) in producing the jth element of the slowness estimate **s**.

With back projection tomography, we do not, in general, have this resolution matrix because we do not reconstruct exactly the generalized inverse (because only a finite number of iterations are applied). We can still make use of $\mathbf{s} = \mathbf{R}\hat{\mathbf{s}}$ for **s** being our model and **R** being the matrix containing the weighting coefficients which scale the actual slowness values to produce the estimated slowness values. It is desirable to have the resolution kernels available for examination. While this is possible with back projection tomography, it is impractical. We instead address resolution by examining the nature of the reconstruction for chosen blocks. By doing this we are examining the amount of model produced by a unit-valued jth

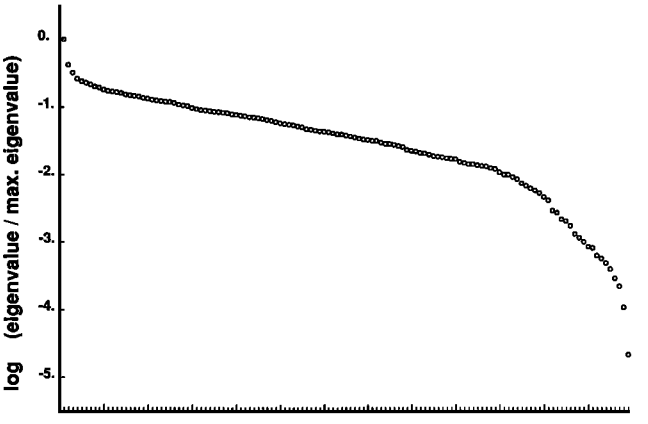

Fig. 4. Plot of $\log_{10}$ (eigenvalue/max eigenvalue) for the 129 significant eigenvalues resulting from the geometry shown in Figure 2. Actual values range from 5.33×10^2 to 1.67×10^{-2} and have been arranged in descending order.

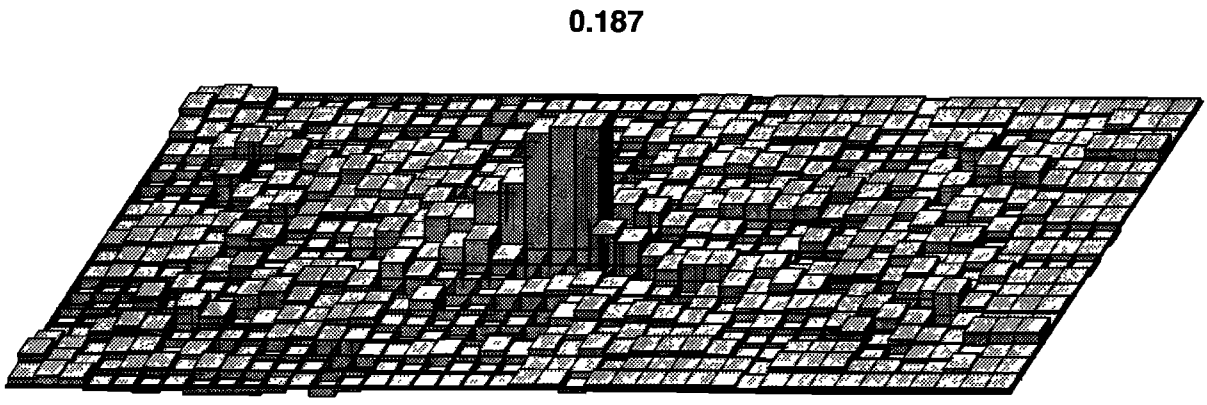

Fig. 5. The resolution kernel for the block located near the center of this figure just to the left of the up-down center line.

block, as opposed to examining the weights of the other model blocks used for the construction of the jth block. Thus while the jth row of $\mathbf{R}$ is the jth resolution kernel, the jth column of $\mathbf{R}$ is identified with the jth spread function. It is noted that when the point spread functions are independent of position and also possess center symmetry, such as is the case with a homogeneous ray set, then $\mathbf{R}$ is symmetrical and the jth point spread function is identical to the jth resolution kernel. When the responses approximately have center symmetry and are similar over the region covered by most of the jth point spread function, then the jth point spread function is very similar to the jth resolution kernel in form. These latter conditions on the point spread function are often the case. When this is so, the observation of a point spread function gives one a direct idea of the corresponding resolution kernels.

If one assumes that all of the estimated variances in the data σ_i^2 are independent, of zero mean, and equal to some constant variance estimate σ^2, the covariance of the model parameters, i.e., covariance of the slowness estimates, $\text{cov}(\mathbf{s}) = \mathbf{s}\mathbf{s}^T$, is commonly used to estimate the effects of noise on the model with $\text{cov}(\mathbf{s}) = \sigma^2(\mathbf{L}^T\mathbf{L})^{-1}$ [e.g., *Jackson*, 1972]. Unfortunately, $(\mathbf{L}^T\mathbf{L})^{-1}$ is not at our disposal. To test the sensitivity of the model to noise, a direct inversion is run on a Gaussian distribution of time delays which are input as though they were data, and the resulting model is examined. If this noise were to have no influence, the model would everywhere be equal to zero. If exceptional sensitivity to noise were found, this procedure lacks the ability to identify explicitly the eigenvectors responsible for the sensitivity or to quantify accurately the degree of sensitivity these eigenvectors have to noise, i.e., determine the eigenvalues. It is a simple matter, however, to determine the variance of the resulting model, which is a general statistical description on the effects of the input noise. This procedure then can test if the inversion is sufficiently insensitive to noise, but if failings occur, it cannot be specific about this, short of an overall statistical description of the failure. Because the noise used for this test is randomly generated, it is possible that the potentially sensitive eigenvectors happen to miss being excited. For this reason, several sets of random noise have been tried, all giving statistically similar results. If any sensitive eigenvectors are present, they have passed all of our tests unexcited.

Numerical Tests of Tomographic Schemes

In this section the procedures outlined above are applied to a synthetic test structure. To do this, a test structure and a ray set are chosen and "observed" delays are calculated. These delays are then inverted. (It is pointed out that the ray paths are assumed and there are therefore no ray-tracing problems such as may exist when inverting real data.) The ray set chosen is shown in Figure 2*a*; the ray distribution is inhomogeneous and strongly anisotropic. There are 263 rays in this set. Figure 2*b* shows the actual structure. There are $19 \times 40 = 760$ blocks in the model. Notice that the ray set and the actual structure are symmetrical about the horizontal midline, and hence the reconstructions will be similarly symmetric. The test structure was chosen to show the point response in each of the two basic regions represented, one with good angular coverage and one with restricted coverage. Figures displaying the results of the various methods are shown below. These figures are grouped according to type of procedure applied: weighting to compensate for anisotropy, iteration, and filtering. In each figure, only the application of the

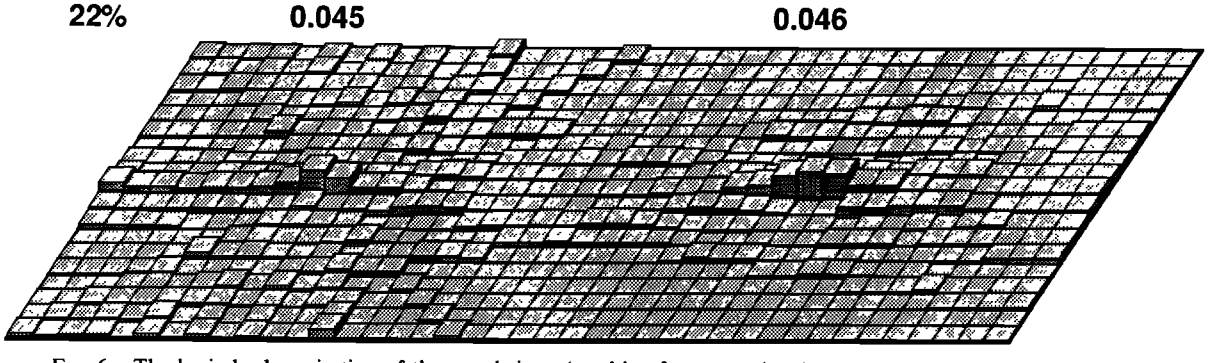

Fig. 6. The basic back projection of the travel times (resulting from equation (1)) produced by the rays and structure show in Figure 2. Notice the low level of reconstruction and unfocused nature of the model.

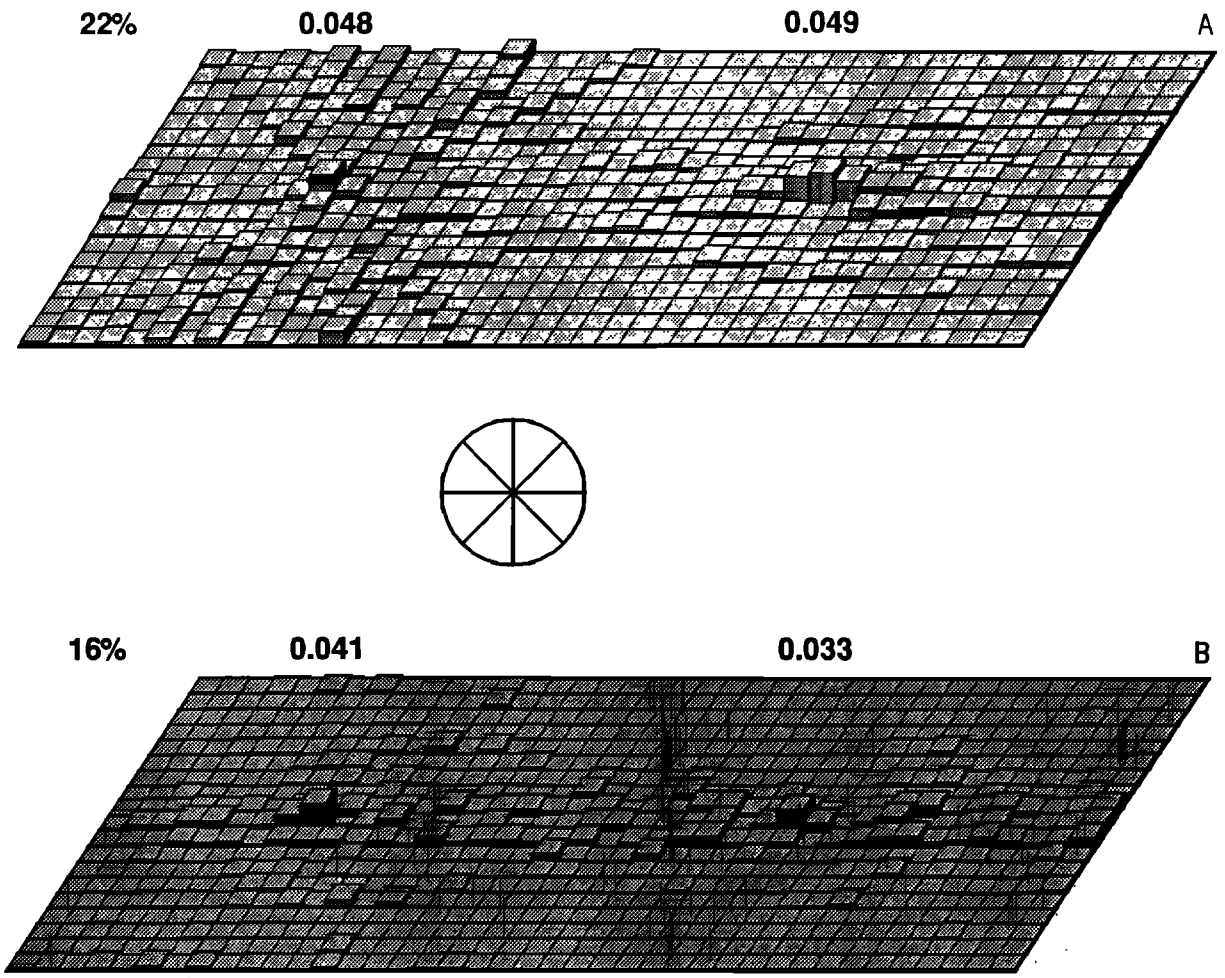

Fig. 7. Graphs showing the effect of the two weighting schemes discussed in the text. (*a*) Made by weighting each ray with a weighting function that gives length-traversing rays 30% as much weight as width-traversing rays. (*b*) Ray weighting is accomplished by binning of the rays according to the orientation angle of the ray, as described in the text. The "pie slice" diagram is a top view of the four orientation bins used, where the two "slices" directly across from one another composes a single orientation bin.

single procedure being investigated is shown in order to isolate its effect. In practice, better results can be obtained by using a combined set of procedures, one from each group, and one such example is shown.

We can gain insight by first studying the generalized inverse solution to this problem. Since the data were created directly from the actual structure, at least this exact solution must exist. Using the symmetry of the problem, the problem can be reduced to $10 \times 40 = 400$ blocks and 141 rays. It should be clear that the problem is underdetermined, and therefore a suite of solutions exist that exactly satisfy the data. The generalized inverse is the solution from this suite with the least energy [*Backus and Gilbert*, 1968]. This model is shown in Figure 3. Notice that this is not the actual structure, i.e., Figure 2*b*, and we can expect to do no better than to reconstruct this solution. The 141 rays result in 129 eigenvalues of $\mathbf{L}^T\mathbf{L}$, indicating a slight redundancy in the data. These eigenvalues, ordered by descending value, are plotted in Figure 4. The maximum eigenvalue of $\mathbf{D}^{-1}\mathbf{L}^T\mathbf{N}^{-1}\mathbf{L}$ is 0.991, and thus the iterative solution using (6) is guaranteed to convergence. Figure 5 shows the resolution kernel for a block located near the model center.

Figure 6 shows the basic back projection inverse given by (1). Notice the overall low amplitude of the model and the tendency of the reconstruction to be more strongly blurred along the paths most commonly taken by rays. Figure 7 shows the effects of the two previously discussed weighting schemes. Reducing the weight of rays that traverse the model in the direction most commonly taken reduces the tendency to streak in that direction. Figure 7*a* shows the result of giving length-traversing rays 30% as much weight as height-traversing rays in (3). When ray directionality is fairly homogeneous, direct ray weighting works well, as seen for the left-side anomaly in this figure. Figure 7*b* shows that the binning approach to weighting adapts itself fairly well to local variations in ray distribution without making explicit use of information about the rays. This suggests that a weighting scheme should be chosen with consideration to the ray geometry one is dealing with.

Next are shown the results of applying iteration. The basic back projection inverse (Figure 6) is the model produced by the first iteration. Using (5) with 5, 50, and 500 iterations results in the models shown in Figures 8*a*, 8*b*, and 8*c*, respectfully. The result of applying 50 iterations with (6) is shown in Figure 9. It is seen that through repeated iteration the model becomes more focused but that in those areas

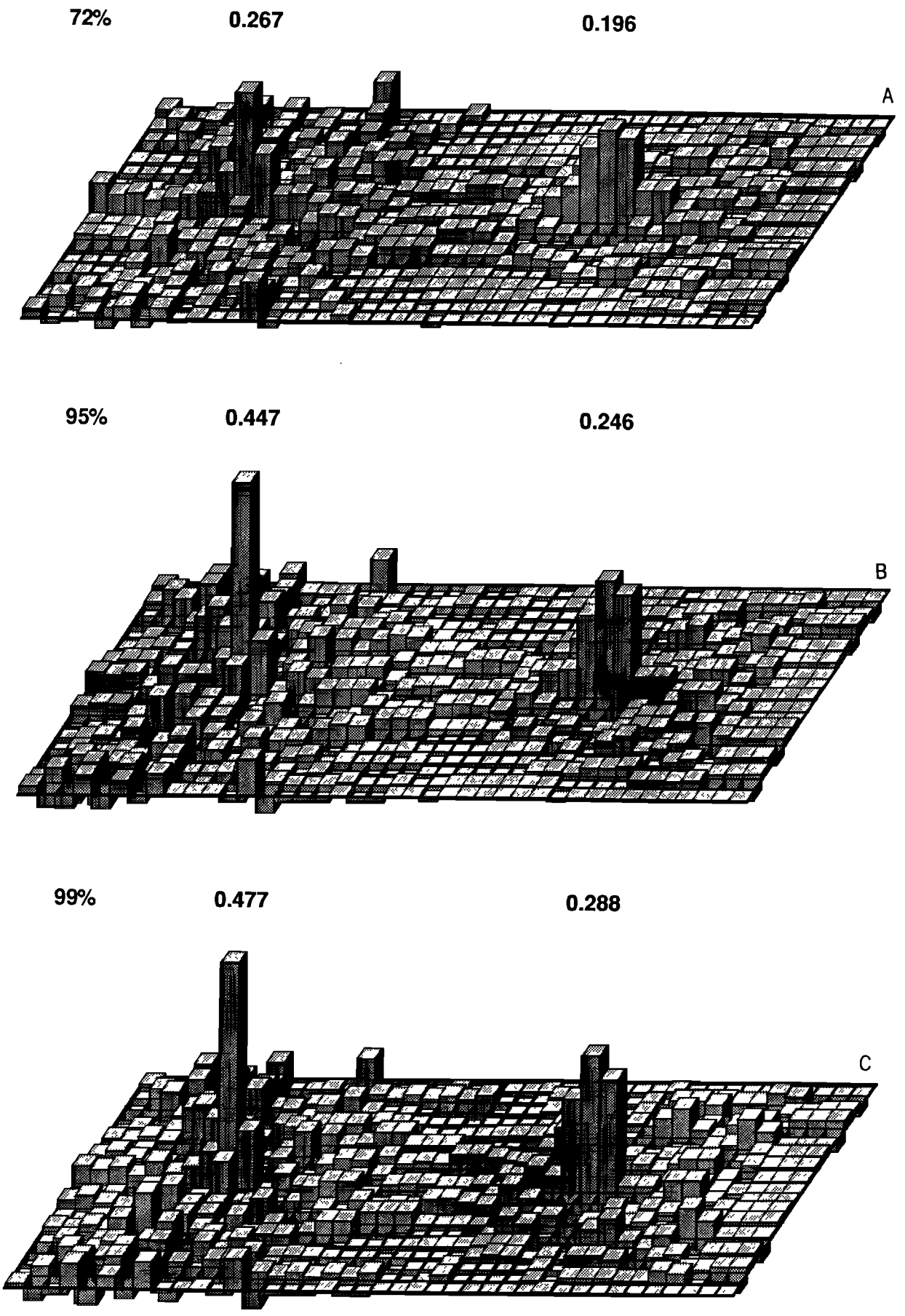

Fig. 8. Models resulting from repeated iteration on the travel time residuals (equation (5)). (*a*) Five iterations have been applied, (*b*) fifty iterations have been applied, and (*c*) five hundred iterations have been applied.

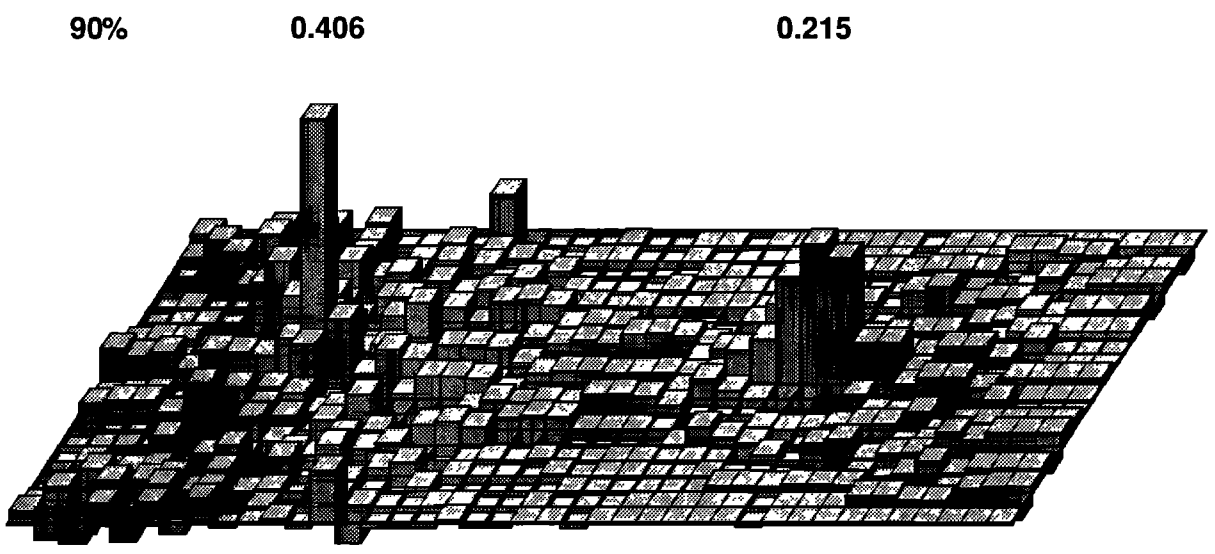

Fig. 9. The model resulting from 50 repeated iterations on the travel time residuals with (6). The results are similar to those produced with (5) (Figure 8*b*), though the degree of reconstruction is not as complete.

where there are few crossing rays the rate of improvement is slow. Figure 10 shows the amount of data unaccounted for as a function of iteration number. The values displayed are $\log_{10}$[rms(residuals)/rms(data)], where rms refers to root-mean-square. Using (5), 1000 iterations yields 99.7% of the data (in an rms sense), while (6) yields 98.8%. In both cases the model converges very nearly to that given by the generalized inverse (Figure 3), though to get very close, very many iterations are needed.

The effect of filtering is illustrated with Figure 11. Figure 11*a* gives the results of applying the Laplacian filter (equation (4*b*)). Figure 11*b* shows the response when the filter is spatially broader. Extending this filter is done to simulate **hd** more closely, and in this case, includes the 7 × 3 rectangular patch centered on the central block and long in the left-right direction. These blocks have been weighted in decreasing amounts as distance increases away from the central block. Figure 11*c* has been produced through wave number domain deconvolution. In this case, an empirical Green's function is constructed by determining the average of the two-point spread functions produced by the separate basic back projection reconstructions of the two anomalous blocks. Wave number domain deconvolution is often unstable, producing an erratic, high wave number rich model. For this reason, various stabilization techniques are used. If one is satisfied with a less focused model, we can attempt to reconstruct a narrow peak instead of a single-block anomaly. This can be done by dividing the wave number domain representation of the Green's function by the wave number domain representation of the desired narrow peak and deconvolving the present model with this function (similar to shaper filtering, e.g., *Tritel and Robinson* [1966]). To further increase stability, the empirical Green's function is tapered to zero at a distance of 6 blocks from the central block. Also, the denominator is clamped so that it never dropped below a value of half the rms value of the set of transform coefficients. The application of this set of stabilizing procedures results in a reconstruction represented only by wave numbers of intermediate value because the high wave numbers have been eliminated with the clamping step, and the lower wave numbers have been removed by the windowing of the Green's function. These efforts yield the model shown in Figure 11*c*. The amount of inhomogeneity in the ray set is greater than is well suited for deconvolution, and the resulting model is not as good as that produced with space domain filtering (Figures 11*a*–11*b*). (This is not always the case. Teleseismic arrivals to a seismic array, for instance, typically have a sufficiently homogeneous ray distribution to allow this method to work well [e.g., *Humphreys et al.*, 1984].)

Finally, Figure 12 shows a reconstruction produced with the combined application of several techniques. In the construction of this model the binning approach to weighting is used, filtering is with the broad filter, and 15 iterations are applied. This reconstruction is fairly good considering that only 15 iterations have been applied. Other combinations of techniques or a greater number of iterations are, of course, possible.

We now consider resolution and the effects of noise. As mentioned above, with back projection tomography, resolution is addressed not with "resolution kernels" [*Wiggens*, 1972] but rather with point spread functions. Figure 13 shows the point spread function for a central point. In this case, the point spread function has been generated with (5) and 50 iterations and thus corresponds to Figure 8*b*. Figure 5 shows the corresponding resolution kernel for this point. These two func-

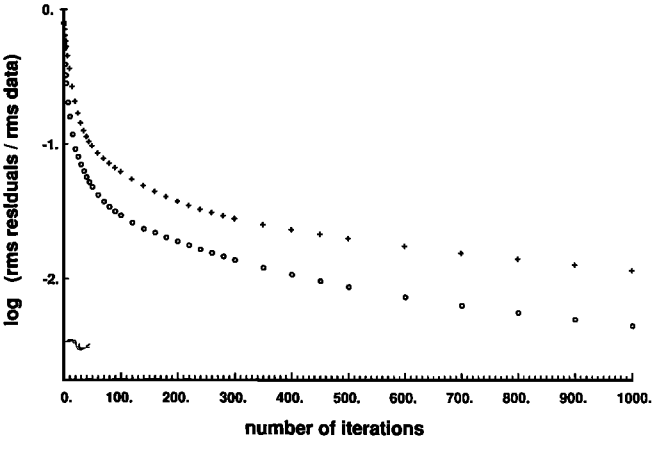

Fig. 10. Plots showing convergence as a function of iteration. Shown is $\log_{10}$[rms(residuals)/rms(data)] as a function of iteration number. The circles represent the application of (5), while the plus symbols represent the application of (6).

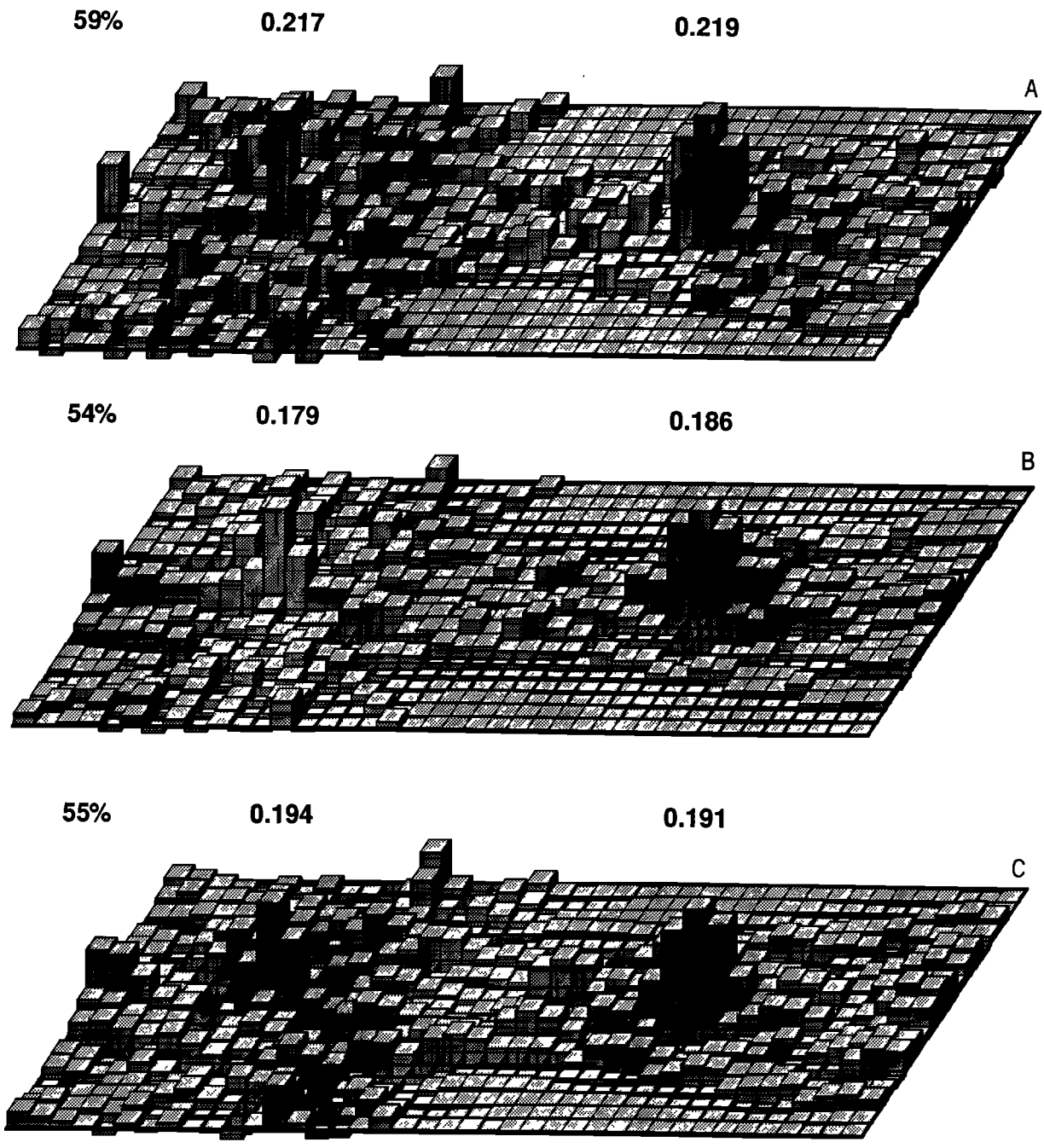

Fig. 11. Graphs showing the effectiveness of filtering. (*a*) The nearest-neighbor filter has been applied to the basic back projection (Figure 6), while (*b*) was made with a broader filter. Figure 11*c* shows the effect of wave number domain deconvolution. A complete discussion is given in the text.

tions are very similar even though the ray geometry is fairly inhomogeneous, indicating the utility of the easy to generate point spread function as a substitute for the resolution kernel.

Figure 14 shows the result of a test run on randomly generated noise, where the same set of procedures that produced Figure 12 were applied. The noise is a Gaussian set of delays with standard deviation equal to half the delay experienced by a ray passing through one of the anomalous blocks. Several such noise tests were run, and all gave statistically similar results. The random, low values of the noise-produced model suggest that the inversions should be well behaved in the presence of noisy data. To show this explicitly, test cases have been run with data created synthetically as before, but now random noise has been added. Figure 15 shows the result of such an inversion as well as a least squares inversion resulting from the same data. Notice that both models are symmetrical about the lengthwise midline. This condition has been imposed and is not normal when random noise is included. As discussed above, the symmetry inherent in the noise free problem has been used to reduce the comutational demands of the least squares inversion. The symmetry shown in the noisy case is a consequence of using these programs. The tomographic results have been made symmetrical (for this figure) in the same way so that direct comparison can be made. The noise is generated in the same manner as that in Figure 14, but, because of the symmetry, only 141/263 of the rays have been used. Statistically this is equivalent to increasing the noise level by a factor of 1.37. This "data set" has 10 times as much

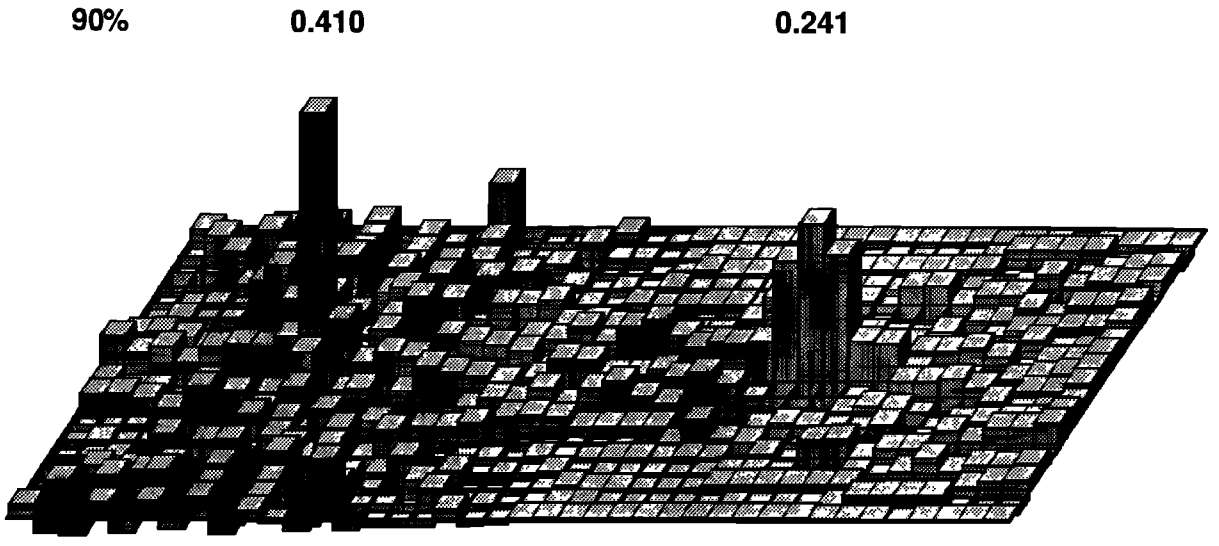

Fig. 12. An example of the integrated application of techniques. Binning has been used for ray weighting, filtering was applied in the space domain with a broad filter, and 15 iterations were applied. Other combinations of techniques and the application of more iterations are also possible.

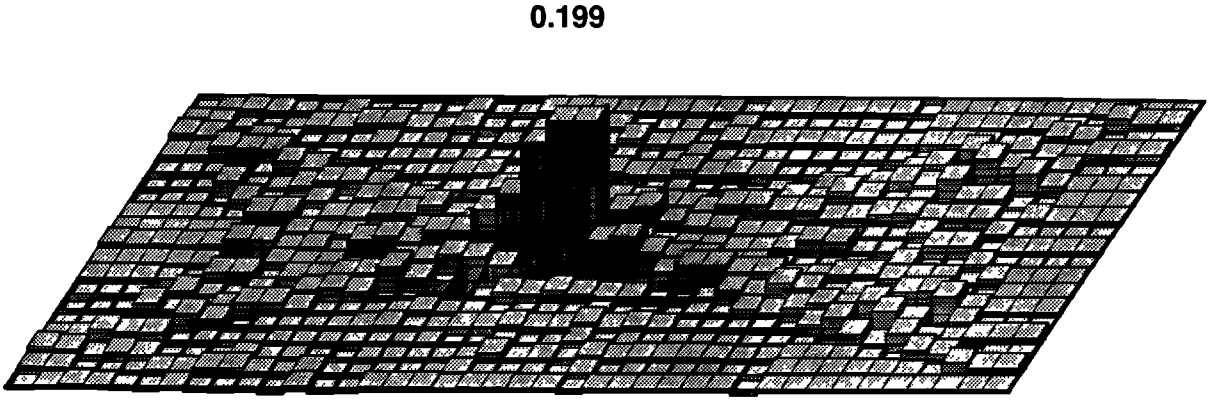

Fig. 13. The point spread function of the same block whose resolution kernel is shown in Figure 5. Notice the similarity of these two functions (in spite of the ray inhomogeneity), indicating the utility of the point spread functions for representing the nature of resolution.

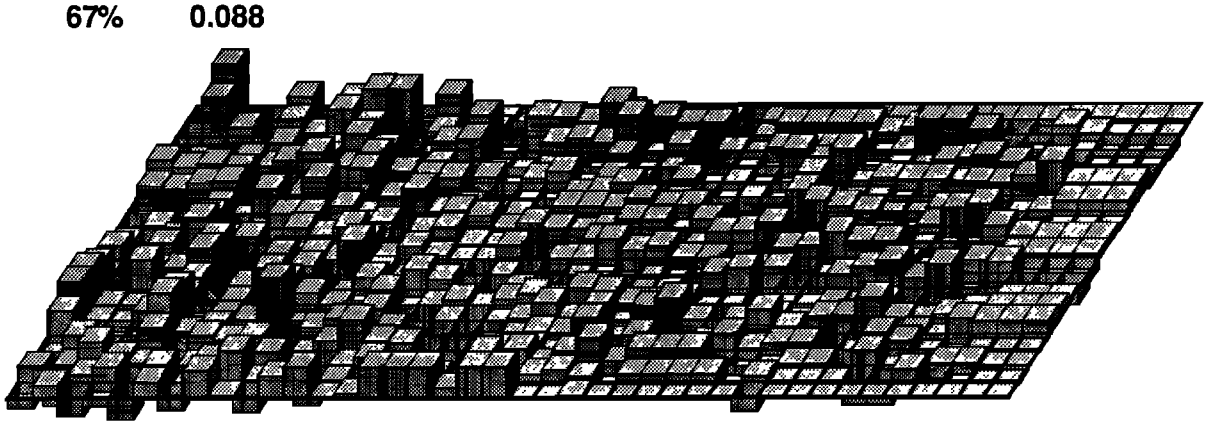

Fig. 14. The model resulting from the inversion of random noise. This graph was created by applying identically the same procedure which were used in the generation of Figure 12. The time delays are Gaussian and have 10 times the signal as that in the data resulting from Figure 2.

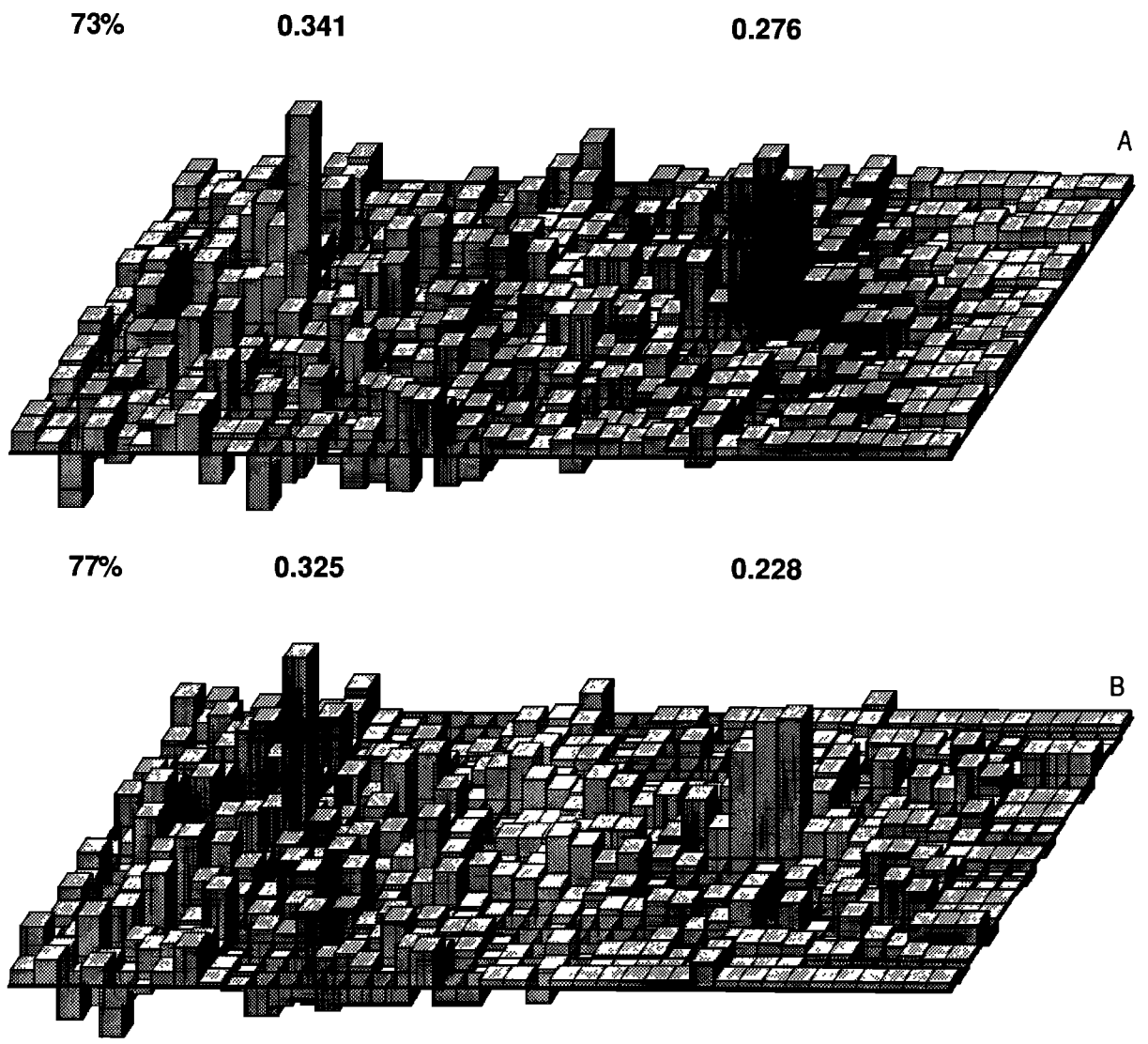

Fig. 15. The model resulting from time delays containing both random noise and coherent signal. These delays are therefore similar to "real" data. The noise is statistically similar to the noise that generated Figure 14, but the average level of noise in Figure 14 is less by a factor of 1.37. (*a*) The inversion procedure is similar to that used in the generation of Figures 12 and 14. In the generation of this model, however, symmetry about the lengthwise midline was imposed so that the figure can be directly compared to Figure 15*b*. (*b*) The same data and noise were inverted with the damped least squares method. A damping constant of 0.2 was found to give the best model, and this is the constant that was used to generate this model.

signal in the random noise as it does resulting from the actual velocity structure.

Figure 15*a* shows a back projection tomography solution to this data set, where we have used the same algorithm that was used to generate Figure 12. Figure 15*b* shows a least squares solution with damping constant equal to 0.2. Several damping coefficients were tried in the least squares inversion, and a constant of 0.2 was chosen as giving the "best" model (based on visual appearance). The results of the two inversions are very similar even with respect to details. The major difference is the overall slightly smaller amplitude of the least squares generated model and choosing a damping constant of slightly less value would eliminate this difference. The ability of the procedure to reconstruct the structure in the presence of a relatively high level of noise attests to the robustness of the inversions.

DISCUSSION

Filtering and iteration both attempt to generate the least squares solution (that is, they approximately solve the normal equations), and we therefore expect that the slowness structure determined by these methods to be quite similar. The manner in which the various methods reconstruct an image differ, however, and the differing characteristics of each can often be taken advantage of. Iteration focuses an image through repeated back projection of the existing residuals, and as such, it is a very stable procedure. Furthermore, it has no inherent dependency upon the ray geometry. A desirable aspect of iteration is that the iterative sequence can be stopped at the point where the structure is resolved to the degree one feels is warranted by the data. This takes advantage of the back projection's inherent smoothing properties and is somewhat analogous to diminishing the influence of the small eigenvectors of

$\mathbf{L}^T\mathbf{L}$ when inverting with the generalized inverse [*Ivansson*, 1983]. The major problem with iteration is that the rate at which the reconstruction becomes focused can be very slow, especially along corridors where there are few crossing rays. Ray weighting tends to lessen the severity of this problem. Filtering, on the other hand, is a relatively rapid process. When the ray geometry is nearly homogeneous, this approach works very well. This is true even if the ray geometry is strongly anisotropic. The major drawback to filtering is that it can produce a rather unstable model when used with a ray distribution that is inhomogeneous.

An alternative inversion technique is the generalized inverse [e.g., *Backus and Gilbert*, 1968]. This method is straightforward and well understood, and a description of resolution and noise sensitivity are natural products of the inversion. However, the inversion requires the explicit construction and inversion of an $M \times M$ matrix, where M is the number of model parameters used to describe the slowness structure (usually, the number of blocks used). Because the storage and inversion of large matrices is a computer intensive process, one is often forced to limit the number of model parameters. The resolution of such a model may therefore be limited not by the inherent restrictions of the data but rather by the coarseness of the model chosen. The computer requirements for the construction of the figures shown above are given as an example: By making use of the symmetry in the ray geometry and in the actual structure during the least squares inversions, memory requirements were cut by a factor of 4, and CPU time was cut by a factor of 8. Even so, solutions required 4 Mbytes of memory and $3\frac{1}{2}$ CPU hours on a Masscomp 5400, as compared to 25 kbytes of memory and 1-CPU min for the tomographic solution shown in Figure 12 (where no regard was paid to symmetry properties and the entire model was constructed).

Acknowledgments. This work has benefited from many discussions with Rob Comer, John Fawcett, and Tom Hearn, a group of people brought together at the California Institute of Technology (Caltech) by their mutual interest in the theory and application of back projection tomography. We would like to thank Bob Langan and anonymous reviewers who contributed to the quality of this paper. This study was supported by the U.S. Geological Survey (USGS 14-08-0001-G1171) and by an NSF-PYI award to RWC. Contribution 4502 from the Seismological Laboratory, Caltech.

REFERENCES

Aki, K., and P. Richards, *Quantitative Seismology: Theory and Methods*, 932 pp., W. H. Freeman, New York, 1980.

Aki, K., A. Christoffersson, and E. S. Huseby, Determination of three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, *82*, 277–296, 1977.

Backus, G., and F. Gilbert, The resolving power of gross earth data, *Geophys. J. R. Astron. Soc.*, *16*, 169–205, 1968.

Bracewell, R. N., Strip integration in radio astronomy, *Aust. J. Phys.*, *9*, 198–217, 1956.

Bracewell, R. N., *The Fourier Transform and its Applications*, p. 253, McGraw-Hill, New York, 1965.

Dines, K. A., and R. J. Lytle, Computerized geophysical tomography, *Proc. IEEE*, *67*, 1065–1073, 1979.

Gilbert, P. F. C., Iterative methods for three-dimensional reconstructions of an object from projections, *J. Theor. Biol.*, *36*, 105–117, 1972.

Herman, G. T., A. Lent, and S. W. Rowland, ART: Mathematics and Applications, *J. Theor. Biol.*, *42*, 1–32, 1973.

Humphreys, E., R. W. Clayton, and B. H. Hager, A tomographic image of mantle structure beneath southern California, *Geophys. Res. Lett.*, *11*, 625–627, 1984.

Ivansson, S., Remark on an earlier proposed iterative tomographic algorithm, *Geophys. R. J. Astron. Soc.*, *75*, 855–860, 1983.

Jackson, D. D., Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. Astron. Soc.*, *28*, 97–109, 1972.

Natterer, F., *The Mathematics of Computerized Tomography*, 222 pp., John Wiley, New York, 1986.

Nolet, G., Solving or resolving inadequate and noisy tomographic systems, *J. Comput. Phys.*, *61*, 463–482, 1985.

Paige, C. C., and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *Assoc. Comput. Mach. Trans. Math. Software*, *8*, 43–71, 1982*a*.

Paige, C. C., and M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *Assoc. Comput. Mach. Trans. Math. Software*, *8*, 195–209, 1982*b*.

Radon, J., Uber die Bestimmung von Funktionen durch ihre Integralwerte langs gewisser Mannigfaltigkeiten, *Ber. Sachsische Akad. Wiss. Leipzig, Math. Phys. Kl.*, *69*, 262–267, 1917.

Rowland, S. W., Computer implementation of image reconstruction formulas, in *Image Reconstructions from Projections*, edited by G. T. Herman, Springer-Verlag, New York, 1979.

Tritel, S., and E. A. Robinson, The design of high-resolution digital filters, *IEEE Trans. Geosci. Electron.*, *4*, 25–38, 1966.

Wiggins, R., The general linear inverse problem: Implications of surface waves and free oscillations on earth structure, *Rev. Geophys.*, *10*, 251–285, 1972.

R. W. Clayton, Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125.

E. Humphreys, Department of Geological Sciences, University of Oregon, Eugene, OR 97403.

(Received December 29, 1986;
revised August 4, 1987;
accepted November 19, 1987.)